Universidad De Sonora
Ciencias Exactas y Naturales

Inteligenia artificial

Evaluación de competencias:

"Introducción: fundations"

Autor:

Luis Juventino Velazquez Hidalgo

1 Optimization and Probability

a. Let $\mathbf{x_1}, ..., \mathbf{x_n}$ be real numbers representing positions on a number line. Let $\mathbf{w_1}, ..., \mathbf{w_n}$be positive real numbers representing the importance of each of these positions. Consider the quadratic function: $\mathbf{f}(\theta) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{w_i}(\theta - \mathbf{x_i})^{\mathbf{2}}$. Note that θ here is a scalar. What value of θ minimizes $\mathbf{f}(\theta)$? Show that the optimum you find is indeed a minimum. What problematic issues could arise if some of the $\mathbf{w_i}$'s are negative?

Note: You can think about this problem as trying to find the point θ that's not too far away from the $\mathbf{x_i}$'s. Over time, hopefully you'll appreciate how nice quadratic functions are to minimize.

What we expect: An expression for the value of θ that minimizes $\mathbf{f}(\theta)$ and how you got it. A short calculation/argument to show that it is a minimum. 1-2 sentences describing a problem that could arise if some of the $\mathbf{w_i}$'s are negative.

b. In this class, there will be a lot of sums and maxes. Let's see what happens if we switch the order. Let $\mathbf{f}(\mathbf{x}) = \max_{\mathbf{s}\in|-\mathbf{1},\mathbf{1}|} \sum_{\mathbf{i=1}}^{\mathbf{d}} \mathbf{s}\mathbf{x_i}$ and $\mathbf{g}(\mathbf{x}) = \sum_{\mathbf{i=1}}^{\mathbf{d}} \max_{\mathbf{s}\in|-\mathbf{1},\mathbf{1}|} \mathbf{s_i}\mathbf{x_i}$, where $\mathbf{x} = (\mathbf{x_1}, ..., \mathbf{x_d}) \in \mathbb{R}^{\mathbf{d}}$ is a real vector. Which of $\mathbf{f}(\mathbf{x}) \leq \mathbf{g}(\mathbf{x})$, $\mathbf{f}(\mathbf{x}) = \mathbf{g}(\mathbf{x})$, or$\mathbf{f}(\mathbf{x}) \geq \mathbf{g}(\mathbf{x})$ is true for all $\mathbf{x}$? Prove it.

Hint: You may find it helpful to refactor the expressions so that they are maximizing the same quantity over different sized sets.

What we expect: A short (3-5) line/sentence proof. You should use mathematical notation in your proof, but can also make your argument in words.

Respuesta:

$$\mathbf{f}(\mathbf{x}) = \max_{\mathbf{s}\in|-\mathbf{1},\mathbf{1}|} \sum_{\mathbf{i=1}}^{\mathbf{d}} \mathbf{s}\mathbf{x_i} = \max_{\mathbf{s}\in|-\mathbf{1},\mathbf{1}|} \mathbf{s} \sum_{\mathbf{i=1}}^{\mathbf{d}} \mathbf{x_i} = \max(\sum_{\mathbf{i=1}}^{\mathbf{d}} \mathbf{x_i}, -\sum_{\mathbf{i=1}}^{\mathbf{d}} \mathbf{x_i}) = |\sum_{\mathbf{i=1}}^{\mathbf{d}} \mathbf{x_i}|$$

De una forma identica ...

$$\mathbf{g}(\mathbf{x}) = \sum_{\mathbf{i=1}}^{\mathbf{d}} \max_{\mathbf{s}\in|-\mathbf{1},\mathbf{1}|} \mathbf{s_i}\mathbf{x_i} = \mathbf{g}(\mathbf{x}) = \sum_{\mathbf{i=1}}^{\mathbf{d}} |\mathbf{x_i}|$$

Es oportuno mensionar que la suma del valor aboluto de x_i siempre va a ser **mayor o igual** que el valor absoluto de la suma de x_i

$$\therefore f(x) \leq g(x)$$

c. Suppose you repeatedly roll a fair six-sided die until you roll a **1** or a **2** (and then you stop). Every time you roll a **3**, you lose **a** points, and every time you roll a **6**, you win **b** points. You do not win or lose any points if you roll a **4** or a **5**. What is the expected number of points (as a function of **a** and **b**) you will have when you stop?

Hint: You will find it helpful to define a recurrence. If you define **v** as the expected number of points you get from playing the game, what happens if you roll a **3**? You lose **a** points

and then get to play again. What about the other cases? Can you write this as a recurrence?

What we expect: A recurrence to represent the problem and the resulting expression from solving the recurrence (no more than 1-2 lines).

Respuesta: Si tomamos el consejo y definimos v como una funcion recursiva que representa el valor esperado, entonces tenemos que:

$$v = \frac{1}{3}(v) + \frac{1}{6}(v-a) + \frac{1}{6}v(+b) + \frac{1}{3}(v*0)$$

Donde cada fraccion es el valor esperado de optener esa cantidad de puntos y el $(v*0)$ representa que el juego se ha acabado (Reseteando los puntos a 0)

d. Suppose the probability of a coin turning up heads is $\mathbf{0 < p < 1}$, and we flip it **6** times and get $\{\mathbf{T}, \mathbf{H}, \mathbf{H}, \mathbf{H}, \mathbf{T}, \mathbf{H}\}$. We know the probability (likelihood) of obtaining this sequence is $\mathbf{L(p) = (1-p)ppp(1-p)p = p^4(1-p)^2}$. What value of $\mathbf{p}$ maximizes $\mathbf{L(p)}$? Prove/Show that this value of $\mathbf{p}$ maximizes $\mathbf{L(p)}$. What is an intuitive interpretation of this value of $\mathbf{p}$?

Hint: Consider taking the derivative of $\log \mathbf{L(p)}$. You can also directly take the derivative of $\mathbf{L(p)}$, but it is cleaner and more natural to differentiate $\log \mathbf{L(p)}$. You can verify for yourself that the value of $\mathbf{p}$ which maximizes $\log \mathbf{L(p)}$ must also maximize $\mathbf{L(p)}$ (you are not required to prove this in your solution).

What we expect: The value of $\mathbf{p}$ that maximizes $\mathbf{L(p)}$ and the work/calculation used to solve for it. Note that you must prove/show that it is a maximum. A 1-sentence intuitive interpretation of the value of $\mathbf{p}$.

Respuesta: Aplicaremos la derivada logaritmica:

$$\ln(y) = \ln(\left((1-p)^2 p^4\right) : \frac{y'}{y} = \frac{6p-4}{p^2-p} : y' = \frac{6p-4}{p^2-p}(\left((1-p)^2 p^4\right)$$

Encontrando el minimo :

$$\frac{6p-4}{p^2-p}(\left((1-p)^2 p^4\right) = 0 : \frac{6p-4}{p^2-p} = 0 := 6p-4 = 0 := p = \frac{4}{6} = \frac{2}{3}$$

Encontraremos si es minimo o maximo, ahora usaremos derivada directa:

$$: y'' = \frac{6p-4}{p^2-p}(\left((1-p)^2 p^4\right) = -\frac{2\left(p-1\right)^2 p^4 \left(3p^2-4p+2\right)}{\left(p-1\right)^2 p^2}$$

Con $y(p)''$ con $p = \frac{2}{3}$ ... El dividendo es un número positivo multiplicado por un otro número positivo, da positivo. En el dividendo, los tres primeros terminos son positivos y $3(\frac{2}{3})^2 - 4(\frac{2}{3}+2) = \frac{12}{9} - \frac{8}{3} + 2$, podemos ver que el ultimo termino tambien es positivo, por lo que la division da como resultado un número positivo. Multiplicando la formula por la negacion, la expresion nos da un número negativo $\therefore$ El punto de inflexion es un maximo

Esto se interpreta como que: "Si $p = \frac{2}{3}$, nuestra probabilidad de optener $\{\mathbf{T}, \mathbf{H}, \mathbf{H}, \mathbf{H}, \mathbf{T}, \mathbf{H}\}$ es maxima"

e. Now for a little bit of practice manipulating conditional probabilities. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are two events such that $\mathbf{P(A|B) = P(B|A)}$. We also know that and $\mathbf{P(A \cup B) = \frac{1}{2}}$ and $\mathbf{P(A \cap B) > 0}$. Prove that $\mathbf{P(A) > 1/4}$.

Hint: Note that $\mathbf{A}$ and $\mathbf{B}$ are not necessarily mutually exclusive. Consider how we can relate $\mathbf{P(A \cup B)}$ and $\mathbf{P(A \cap B)}$.

What we expect:A short (∼5 line) proof/derivation.

Respuesta: El problema nos dice que:

$$P(A \mid B) = P(B \mid A) \equiv \frac{P(A \bigcap B)}{P(B)} = \frac{P(B \bigcap A)}{P(A)}$$

Y ya que $P(B \bigcap A)$ es un numero mayor a cero :

$$P(B) = P(A)$$

Ahora sigue:

$$P(A \bigcup B) = \frac{1}{2} = P(A) + P(B) - P(A \bigcap B) \equiv$$

$$\frac{1}{2} + P(A \bigcap B) = P(A) + P(B) \equiv \frac{1}{2} > 2P(A) \equiv P(A) > \frac{1}{4}$$

$$\therefore P(A) > \frac{1}{4}$$

f. Let's practice taking gradients, which is a key operation for being able to optimize continuous functions. For $\mathbf{w} \in \mathbb{R}^{\mathbf{d}}$ (represented as a column vector), and constants $\mathbf{a_i}, \mathbf{b_j} \epsilon \mathbb{R}^{\mathbf{d}}$ (also represented as column vectors), $\lambda \epsilon \mathbb{R}$, and a positive integer $\mathbf{n}$, define the scalar-valued function

$$f(w) = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} (a_i^T w - b_i^T w)^2 \right) + \frac{\lambda}{2} ||w||_2^2$$

where the vector is $\mathbf{w} = (\mathbf{w_1}, \ldots, \mathbf{w_d})^{\mathbf{T}}$ and $||\mathbf{w}||_{\mathbf{2}} = \sqrt{\sum_{\mathbf{k=1}}^{\mathbf{d}} \mathbf{w_k^2}} = \sqrt{\mathbf{w^T w}}$ is known as the **L2** norm. Compute the gradient $\nabla \mathbf{f}(\mathbf{w})$.

Recall: the gradient is a d-dimensional vector of the partial derivatives with respect to each $\mathbf{w_i}$:

$$\nabla f(w) \left(\frac{\partial(w)}{\partial w_1}, \ldots, \frac{\partial(w)}{\partial w_d} \right)$$

If you're not comfortable with vector calculus, first warm up by working out this problem using scalars in place of vectors and derivatives in place of gradients. Not everything for scalars goes through for vectors, but the two should at least be consistent with each other (when $\mathbf{d = 1}$). Do not write out summations over dimensions, because that gets tedious.

What we expect: An expression for the gradient and the work used to derive it. ($\sim$ 5 lines). No need to expand out terms unnecessarily; try to write the final answer compactly.

2 Problem 2: Complexity

a: Suppose we have an $\mathbf{n} \times \mathbf{n}$ grid, where we'd like to place 4 arbitrary axis-aligned rectangles (i.e., the sides of the rectangle are parallel to the axes). There are no constraints on the location or size of the rectangles. For example, it is possible for all four corners of a single rectangle to be the same point (resulting in a rectangle of size 0) or for all 4 rectangles to be on top of each other. How many possible ways are there to place 4 rectangles on the grid? In general, we only care about asymptotic complexity, so give your answer in the form of $\mathbf{O(n^c)}$ or $\mathbf{O(c^n)}$ for some integer $\mathbf{c}$.
Note: It is unnecessary to consider whether order matters in this problem, since we are asking for asymptotic complexity. You are free to assume either in your solution, as it doesn't change the final answer.
What we expect:A big-O bound for the number of possible ways to place 4 rectangles and some simple explanation/reasoning for the answer ($\sim$ 2 sentences). **Respuesta:** Un rectangulo se puede representar como un par de puntos donde un punto representa la esquina superior derecha y el otro representa la esquina inferior izquierda. Entonces calcular los posibles rectangulos equivale a calcular el numero de permutaciones de los rectangulos. Ya que los rectangulos se pueden solapar, entonces el número que nos piden son las permutaciones de un rectangulo elevado a la cuarta potencia. Puede esté cometiendo un error al calcular las permutaciones de los demas rectangulos (o puede que se necesiten combinaciones en vez de permutaciones) , pero es no importa al final puesto que el resultado lo da una formula; y por ende el algoritmo es de complejidad constante

$$\therefore big - O = O(1)$$

b: Suppose we have an $\mathbf{n} \times \mathbf{3n}$ grid. We start in the upper-left corner (position $(\mathbf{1}, \mathbf{1})$), and we would like to reach the lower-right corner (position $(\mathbf{n}, \mathbf{3n})$) by taking single steps down or to the right. Suppose we are provided with a function $\mathbf{c(i, j)}$ that outputs the cost of going through position $(\mathbf{i}, \mathbf{j})$, and assume it takes constant time to compute for each position. Note that $\mathbf{c(i, j)}$ can be negative. Give an algorithm for computing the cost of the minimum-cost path from $(\mathbf{1}, \mathbf{1})$ to $(\mathbf{n}, \mathbf{3n})$ in the most efficient way (with the smallest big-O time complexity). What is the runtime (just give the big-O)?
What we expect: A description of the algorithm for computing the cost of the minimum-cost path as efficiently as possible ($\sim$5 sentences). The big-O runtime and a short explanation of how it arises from the algorithm. **Respuesta:** Para resolver esté problema usaremos un enfoque de programacion dinamica. Ya que solo se pueden dar pasos a la derecha o a abajo por lo que si tenemos una matriz "**DP**" de dimenciones $[\mathbf{n}][\mathbf{3n}]$ entonces el valor de la casilla $\mathbf{DP[i][1]}$ es :

$$\mathbf{DP[i-1][1] + c(i, j)}$$

Y de forma analoga

$$\mathbf{DP[1][j] = DP[1][j-1] + c(1, j)}$$

Y ya que solo se pude llegar a una casilla desde arriba o desde la izquierda, $\mathbf{DP[i][j]}$ es:

$$\mathbf{c(i, j) + MIN(DP[i-1][j], DP[i][j-1])}$$

Si iniciamos la casilla $\mathbf{DP[1][1] = 0}$ entonces nuestro algoritmo solo tendria que recorrer cada casilla una sola vez

$$\therefore Big - O = O(3n^2)$$

3 Problem 3: Ethical Issue Spotting

One of the goals of this course is to teach you how to tackle real-world problems with tools from AI. But real-world problems have real-world consequences. Along with technical skills, an important skill every practitioner of AI needs to develop is an awareness of the ethical issues associated with AI. The purpose of this exercise is to practice spotting potential ethical concerns in applications of AI - even seemingly innocuous ones.

In this question, you will explore the ethics of four different real-world scenarios using the ethics guidelines produced by a machine learning research venue, the NeurIPS conference. The NeurIPS Ethical Guidelines list sixteen non-exhaustive concerns under Potential Negative Social Impacts and General Ethical Conduct (the numbered lists). For each scenario, you will write a potential negative impacts statement. To do so, you will first determine if the algorithm / dataset / technique could have a potential negative social impact or violate general ethical conduct (again, the sixteen numbered items taken from the NeurIPS Ethical Guidelines page). If the scenario does violate ethical conduct or has potential negative social impacts, list one concern it violates and justify why you think that concern applies to the scenario. If you do not think the scenario has an ethical concern, explain how you came to that decision. Unlike earlier problems in the homework there are many possible good answers. If you can justify your answer, then you should feel confident that you have answered the question well.

Each of the scenarios is drawn from a real AI research paper. The ethics of AI research closely mirror the potential real-world consequences of deploying AI, and the lessons you'll draw from this exercise will certainly be applicable to deploying AI at scale. As a note, you are not required to read the original papers, but we have linked to them in case they might be useful. Furthermore, you are welcome to respond to anything in the linked article that's not mentioned in the written scenario, but the scenarios as described here should provide enough detail to find at least one concern.

What we expect: A 2-5 sentence paragraph for each of the scenarios where you either A. identify at least one ethical concern from the NeurIPS Ethical Guidelines and justify why you think it applies, or B. state that you don't think a concern exists and justify why that's the case. Chosen scenarios may have anywhere from zero to multiple concerns that match, but you are only required to pick one concern (if it exists) and justify your decision accordingly. Furthermore, copy out and underline the ethical checklist item to which you are referring as part of your answer (i.e.: Severely damage the environment). We have also included a citation in the example solution below, but you are not required to add citations to your response.

Example Solution: You work for a U.S. hospital that has recently implemented a new intervention program that enrolls at-risk patients in programs to help address their chronic medical issues proactively before the patients end up in the hospital. The intervention program automatically identifies at-risk patients by predicting patients' risk scores, which are measured in terms of healthcare costs. However, you notice that for a given risk score tier, the Black patients are

considerably sicker when enrolled than white patients, even though their assigned illness risk score is identical. You manually re-assign patients' risk scores based on their current symptoms and notice that the percentage of Black patients who would be enrolled has increased from 17

a: An investment firm develops a simple machine learning model to predict whether an individual is likely to default on a loan from a variety of factors, including location, age, credit score, and public record. After looking through their results, you find that the model predicts mainly based on location and that the model mainly accepts loans from urban centers and denies loans from rural applicants [3]. Furthermore, looking at the gender and ethnicity of the applicants, you find that the model has a significantly higher false positive rate for Black and male applicants than for other groups. In a false positive prediction, a model misclassifies someone who does not default as likely to default.

Respuesta: Para empezar, el algoritmo resulto ser racista. Para calcular el riesgo toma en cuenta genero y la etnia de la persona, los cuales a mi parecer no tiene nada que ver con tu capacidad economica, pero es información sencible que se está usando para denigrar a las personas.

b: Stylometry is a way of predicting the author of contested or anonymous text by analyzing the writing patterns in the anonymous text and other texts written by the potential authors. Recently, highly accurate machine learning algorithms have been developed for this task. While these models are typically used to analyze historical documents and literature, they could be used for deanonymizing a wide range of texts, including code [4].

Respuesta: En el texto mismo se nos dice que podria ser usado para otros propositos menos eticos, así que entrada se esta violando la norma de que
se puede extender a formas dañinas de vigilancia. Y si se ha de extender con esté proposito, estaria violando los derechos humanos, puesto que la privacidad es un derecho y; de principio, si un texto es anonimo es porque el autor no quiere que se le reconozca como suyo

c: A research group scraped millions of faces of celebrities off of Google images to develop facial recognition technology [5]. The celebrities did not give permission for their images to be used in the dataset and many of the images are copyrighted. For copyrighted photos, the dataset provides URL links to the original image along with bounding boxes for the face.

Respuesta: No hay mucho que decir, un principio basico al recolectar la informacion (Además de ser un principio etico general) es que las personas deben dar su consentimiento y que de la misma forma se les comunique el proposito con el cual se están recabando los datos

d: Researchers have recently created a machine learning model that can predict plant species automatically directly from a single photo [6]. The model was trained using photos uploaded to the iNaturalist app by users who consented to use of their photos for research purposes, and the model is only used within the app to help users identify plants they might come across in the wild.

Respuesta: Creo que este es un uso etico de las tecnologias, las personas dieron su consentimiento para recolectar los datos y el modelo solo se usa dentro de la aplicacion desarrollada, por lo que podría apostar que no están compartiendo los datos recolectados con ninguna otra empreza

4 Problem 4: Programming

In this problem, you will implement a bunch of short functions. The main purpose of this exercise is to familiarize yourself with Python, but as a bonus, the functions that you will implement will come in handy in subsequent homeworks.

What we expect: Python code implementing the functions provided in submission.py. Try to make your code as clean and simple as possible and be sure to write your answers between the begin answer and end answer comments.

a: Implement find_alphabetically_first_word in submission.py.

b: Implement euclidean_distance in submission.py.

c: Implement mutate_sentences in submission.py.

d Implement sparse_vector_dot_product in submission.py.

e Implement increment_sparse_vector in submission.py.

f Implement find_nonsingleton_words in submission.py.

Respuesta: Las solucionoes estan implementadas en mi repositorio